Glycobiology vol. 24 no. 1 pp. 85–96, 2014
doi:10.1093/glycob/cwt094
Advance Access publication on October 16, 2013

Defining the structural origin of the substrate sequence independence of O-GlcNAcase using a combination of molecular docking and dynamics simulation

Joanne C Martin[2], Elisa Fadda[2,†], Keigo Ito[3], and Robert J Woods[2,3,1]

[2]School of Chemistry, National University of Ireland, University Road, Galway, Ireland and [3]Complex Carbohydrate Research Centre, University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA

Received on September 10, 2013; revised on October 5, 2013; accepted on October 13, 2013

Protein glycosylation with O-linked *N*-acetylglucosamine (O-GlcNAc) is a post-translational modification of serine/threonine residues in nucleocytoplasmic proteins. O-GlcNAc has been shown to play a role in many different cellular processes and O-GlcNAcylation is often found at sites that are also known to be phosphorylated. Unlike phosphorylation, O-GlcNAc levels are regulated by only two enzymes, O-GlcNAc transferase (OGT) and O-GlcNAc hydrolase (O-GlcNAcase or OGA). So far, no obvious consensus sequence has been found for sites of O-GlcNAcylation. Additionally, O-GlcNAcase recognizes and cleaves all O-GlcNAcylated proteins, independent of their sequence. In this work, we generate and analyze five models of O-GlcNAcylated peptides in complex with a bacterial OGA. Each of the five glycopeptides bind to OGA in a similar fashion, with OGA–peptide interactions primarily, but not exclusively, involving the peptide backbone atoms, thus explaining the lack of sensitivity to peptide sequence. Nonetheless, differences in peptide sequences, particularly at the −1 to −4 positions, lead to variations in predicted affinity, consistent with observed experimental variations in enzyme kinetics. The potential exists, therefore, to employ the present analysis to guide the development glycopeptide-specific inhibitors, or conversely, the conversion of OGA into a reagent that could target specific O-GlcNAcylated peptide sequences.

Keywords: β-*N*-Acetylglucosaminidase (O-GlcNAcase) / GLYCAM / Molecular dynamics / O-Linked *N*-acetyl-glucosamine (O-GlcNAc) / Protein glycosylation

Introduction

Since its discovery (Torres and Hart 1984), the post-translational modification of proteins by β-O-linked *N*-acetylglucosamine (O-GlcNAc) on serine or threonine residues has generated significant interest in the field of glycobiology. The modification occurs on nuclear and cytosolic proteins, many of which modulate cell signaling (Hanover et al. 2010; Slawson et al. 2010). It is also thought that deregulation of this modification may be involved in neurodegenerative diseases, cancer and diabetes mellitus (Hanover 2001; Hart et al. 2007). Additionally, O-GlcNAc rapidly cycles on and off proteins on a time scale similar to that of phosphorylation/dephosphorylation, and thus an interplay has been shown to exist between protein phosphorylation and O-GlcNAcylation (Cheng et al. 2000; Comer and Hart 2000; Kamemura et al. 2002; Zeidan and Hart 2010). Unlike phosphorylation, which is controlled by numerous kinases and phosphatases, the addition and removal of O-GlcNAc is controlled by only two enzymes. O-GlcNAc transferase (OGT) is responsible for the transfer of GlcNAc from UDP–GlcNAc to the protein, whereas O-GlcNAcase (OGA) removes it. Despite the importance of this modification, very little is known about the structural relationship between these enzymes and their substrates. Nor is it clear, how only a single hydrolase is capable of removing O-GlcNAc from all proteins, independent of the protein sequence.

To gain an understanding of the molecular interactions responsible for the ability of OGA to tolerate multiple substrate sequences, we modeled the structure of bacterial OGA, *Bacteroides thetaiotaomicron* GH84 (*Bt*OGA) in complex with five O-GlcNAcylated peptides of biological interest. Notably, the OGAs have been grouped together into the same glycoside hydrolase family GH84 (Henrissat and Davies 1997) based on the similarity of their N-terminal sequence domains. Specifically, *Bt*OGA and human OGA (hOGA) share an overall sequence identity of >40% (Dennis et al. 2006) with the same overall 3D structure, as such the details found from the *Bt*OGA model can be applied to hOGA for the development of therapeutic reagents.

Five glycopeptides, 10 residues in length, were built from each of the following O-GlcNAcylated systems: the tumor suppressor protein p53 (O-GlcNAcylated at Ser149 (Yang et al. 2006)), the DNA packaging protein Histone H3 (O-GlcNAcylated Ser10 (Zhang et al. 2011)), a transforming growth factor binding partner TAB1 (O-GlcNAcylated at Ser395 (Pathak et al. 2012)), as well as c-Myc, a regulator of gene transcription (O-GlcNAcylated at Thr58 (Chou et al. 1995)). Additionally, a synthetic O-GlcNAcylated polyalanine (poly-Ala') peptide was included as a reference.

[1]To whom correspondence should be addressed: Tel: +1-706-542-4454; Fax: +1-706-542-4412; e-mail: rwoods@ccrc.uga.edu
[†]Present address: Department of Chemistry, National University of Ireland, Maynooth, Ireland.

It has been shown that OGA uses a substrate-assisted catalytic mechanism, in which the *N*-acetyl group of O-GlcNAc adopts a *syn*-conformation, in which the amide proton eclipses the proton at C2, which aids the sugar to form the oxime transition state (Dennis et al. 2006; Macauley et al. 2005). Several inhibitors of OGA have been reported (Kim et al. 2006; Whitworth et al. 2006; He et al. 2009; Macauley and Vocadlo 2010), and have been co-crystalized with the enzyme. For this study, a model for the conformation of O-GlcNAc in the binding site of *Bt*OGA was generated based on the atomic coordinates of the inhibitor PUGNAc, a transition state mimic and O-GlcNAc analog (Perreira et al. 2006). The peptides were then conjugated to the bound monosaccharide through a combination of molecular docking and molecular dynamics (MD) simulation.

To provide a structural and dynamic interpretation of the observed ability of OGA to deglycosylate a variety of glycopeptides, each of the complexes was subjected to fully solvated MD simulations (350 ns), followed by direct ΔG energy calculations. We have focused on the conformational properties of these complexes, rather than on the mechanism of deglycosylation, as we are particularly interested in the ability of the enzyme to recognize O-GlcNAc in a variety of protein contexts. We compared the conformation of the bound O-GlcNAcylated peptides to a recently solved bacterial OGA homolog from *Clostridium perfringens* (*Cp*OGA) in complex with an eight residue O-GlcNAcylated p53 peptide (PDBid 2YDR) (Schimpl et al. 2012). In addition, p53 (PDBid 1TUP) (Cho et al. 1994) is the only structurally characterized protein where the O-GlcNAcylation site is positioned in a fully defined loop, thus allowing us to further compare our results (Schimpl et al. 2012).

Results

Structural stability of the Bt*OGA–glycopeptide complexes*

Over the course of the 350 ns MD simulations, performed with no restraints on either the protein or glycopeptides, the protein fold remained stable; the Cα root-mean-square deviation (RMSD) of the *Bt*OGA in all cases reached an equilibrium value of between 0.84 and 0.93 Å (c-Myc′: 0.93 Å, histone H3′: 0.84 Å, p53′: 0.88 Å, poly-Ala′: 0.91 Å and TAB1′: 0.84 Å). The glycopeptides in all cases remained bound to the protein, with the GlcNAc adopting equivalent orientations. The ring conformation in all complexes with *Bt*OGA was monitored over the course of the MD simulations. The analysis showed that in GlcNAc-O-methyl (OMe), GlcNAc-p53, GlcNAc-Histone H3 and GlcNAc-TAB1 the ring adopts conformations between $^{1}S_{3}$ and $^{1,4}B$. In GlcNAc-c-Myc, the sugar ring adopts an $^{4}H_{5}$ conformation, while in GlcNAc-poly-Ala′ the ring is more flexible, adopting two conformations $^{4}H_{3}$ (45%) and $^{1}S_{3}$ to $^{1,4}B$ (55%). In the *Cp*OGA–glycopeptide complexes (Schimpl et al. 2012), the ring adopts conformations between $^{1}S_{3}$ and $^{1,4}B$. The Cremer–Pople pucker parameters (Cremer and Pople 1975) and cannonical ring forms are presented in Supplementary data, Table SI and Figure S2. In addition, the dihedral angles of the GlcNAc-Ser linkages (Φ, Ψ and χ) as well as the NAc conformation were monitored throughout the simulations. The analysis showed that the GlcNAc-Ser linkages, in each of the five glycopeptides, populated comparable rotamers, whilst the NAc group remained stable in the *syn*-orientation throughout (Table I). Additionally, the average dihedral values for these angles in the simulated p53′ and TAB1′–glycopeptide complexes were in close agreement with those seen in the crystal structures of *Cp*OGA complexed with these ligands (Table I).

GlcNAc: A comparison of the hydrogen bond interactions between the GlcNAc and OGA residues for GlcNAc-OMe as well as all five glycopeptides is presented in Supplementry data, Table SII. Hydrogen bonds were found between six OGA residues and the glycan. Of those, five involved sidechain interactions (D344, N339, D242, K166 and H433), whilst the remaining was through a backbone interaction involving OGA residue G135. The interactions between the GlcNAc residue and OGA, in both the p53′ and TAB1′—*Bt*OGA glycopeptide complexes, were in good agreement with those seen in the recently reported crystal structures for the corresponding glycopeptides in complex with *Cp*OGA (Schimpl et al. 2012) (Supplementry data, Table SII). The similarity of the crystallographic and modeled complexes provided independent validation of the computational approach.

Poly-Ala′: The hydrogen bond analysis of the *Bt*OGA—poly-Ala′ complex showed that the most stable hydrogen bond (26% occupancy) was formed between the sidechain of OGA residue D243 with the backbone of Ala at the +2 position (Ala (+2)) of the glycopeptide (positional numbering is relative to the site of glycosylation, Figure 1A). Lower occupancy hydrogen bonds were observed between D243 and Ala(+1) position, and OGA Y137 and Ala(−2) (see Table II). It is notable that no long-lived hydrogen bonds were seen between poly-Ala′ and OGA.

c-Myc′: In the case of the c-Myc′ glycopeptide, the most stable intermolecular hydrogen bond (31.5% occupancy) was observed between the backbone of L56(−2) and OGA residue D243, but unlike the case of poly-Ala′, here the interaction involved only the backbone of D243. Weak (16.8% occupancy) backbone-only interactions were also seen between E54(−4) and OGA residue S245. The remaining hydrogen bonds are almost exclusively through peptide backbone interactions and OGA; none of these were present for over 8.1% of the simulation time.

Histone H3′: The strongest intermolecular hydrogen bond (31.1% occupancy) in the histone H3′ glycopeptide–OGA complex was seen between the backbone of G12(+2) of the glycopeptide and the sidechain of OGA residue D243. This interaction and occupancy is analogous to that seen in the case of poly-Ala′ complex. The second strongest hydrogen bond (25% occupancy)

Table I. Average dihedral angles for GlcNAc-Ser/Thr linkages in each of the five glycopeptide–*Bt*OGA complexes

	Φ[a]	Ψ[a]	χ[a]	NAc[c]
c-Myc′	−51.3 (8)	168.3 (9)	−57.1 (8)	−67.2 (7)
Histone H3′	−78.2 (10)	−176.8 (7)	−63.7 (8)	−64.1 (9)
p53′	−73.3 (9)	−174.0 (8)	−57.8 (9)	−59.2 (9)
p53 X-ray[b]	−74.6	173.2	−52.4	−52.7
Poly-Ala′	−74.8 (13)	176.2 (18)	−63.7 (11)	−63.3 (9)
TAB1′	−86.9 (10)	173.7 (8)	−64.3 (9)	−62.1 (8)
TAB1 X-ray[b]	−67.8	165.8	−52.8	−56.5

[a]Φ = O5-C1-Oγ-Cβ, Ψ = C1-Oγ-Cβ-Cα, χ = Oγ-Cβ-Cα-N, in degrees, with standard deviations in parentheses.
[b]From the p53 and TAB1 glycopeptide–*Cp*OGA complexes (Schimpl et al. 2012).
[c]Dihedral angle C1-C2-N-C.

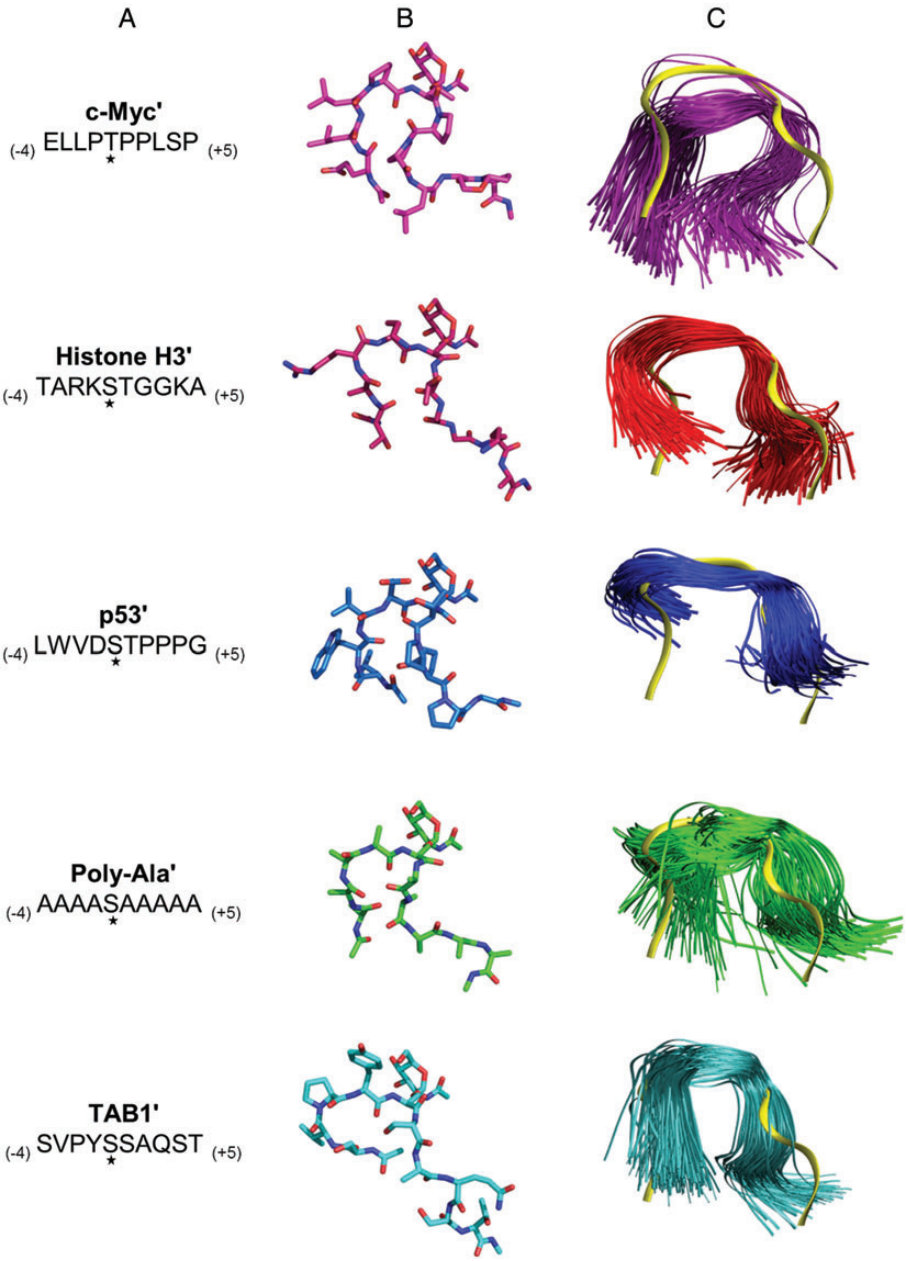

Fig. 1. (A) Five polypeptides were built in extended conformations using the Maestro builder from Schrödinger. Sites of glycosylation are indicated with a star and are positioned at Thr58 (c-Myc), Ser10 (Histone H3), Ser149 (p53), Ser5 (poly-Ala′) and Ser395 (TAB1). **(B)** The top docked conformation of each peptide was used as a starting point for MD. **(C)** Ribbon representation of the peptide structures from the 350 ns simulation aligned onto the GlcNAc residue from the bound p53 glycopeptide – *Cp*OGA crystal structure. The snapshots have been selected every 3.5 ns. The peptides are coloured as follows; c-Myc′ (purple), Histone H3′ (red), p53′ (blue), poly-Ala′ (green), TAB1′ (cyan), with the p53-crystal glycopeptide in yellow.

was observed between the backbone atoms of R8(−2) and the sidechain of OGA residue R347. Additional weak hydrogen bonds were observed between D167, Y137, H433 and D243 of OGA and the peptide; stable for a maximum of 19.2% of the time.

p53′: Glycopeptide p53′ formed a weak intermolecular hydrogen bond (32% occupancy) between the sidechain of W146(−3) and Y550 of OGA. When compared with the recent crystallographic data of the p53 glycopeptide–*Cp*OGA complex (Schimpl et al. 2012), it was found that this hydrogen bond appears to replace the stacking interaction of W146(−3) with protein residue N298 (*Bt*OGA N243). A comparison, of the crystallographic and MD-derived conformations of the p53 glycopeptides with the crystal structure of the full-length p53 protein (PDBid 1TUP) (Cho et al. 1994), indicated that the W146 orientation from the MD results was in good agreement with the conformation seen in the intact p53 protein (Figure 2). The backbone conformation of the MD-derived p53′ glycopeptide gave rise to an average RMSD

Table II. Intermolecular hydrogen bond occupancies between BtOGA and the peptide moieties for each of the five glycopeptide–OGA complexes

Donor H-X	Acceptor	Occupancy
c-Myc′		
L(−2)-Main	D243-Main	31.5
S245-Main	E(−4)-Main	16.8
E(−4)-Main	S245-Main	8.1
H433-Side	T*(0)-Main	7.5
P(+1)-Main	D243-Side	7.1
S245-Side	E(−4)-Side	6.5
Histone H3′		
G(+2)-Main	D243-Side	31.1
R347-Side	R(−2)-Main	25.4
R(−2)-Side	D167-Side	19.2
Y137-Side	S*(0)-Side	13.5
H433-Side	S*(0)-Main	10.9
T(+1)-Main	D243-Side	5.4
p53′		
W(−3)-Side	Y550-Side	31.9
H433-Side	T(+1)-Side	18.3
L(−4)-Main	D243-Main	14.8
Y137-Side	S*(0)-Side	8.3
W(−3)-Side	K166-Main	7.8
V(−2)-Main	Y137-Side	6.7
T(+1)-Side	H433-Side	5.9
Poly-Ala′		
A(+2)-Main	D243-Side	26.0
A(+1)-Main	D243-Side	5.5
A(−2)-Main	Y137-Side	5.4
TAB1′		
A(+2)-Main	D243-Side	24.3
H433-Side	Y(−1)-Side	7.9
Q(+3)-Side	W286-Side	6.9
Y137-Side	S*(0)-Side	6.7

[a]Based on a maximum donor–acceptor distance of 3.5 Å with a minimum occupancy of 5%.

of 1.4 Å relative to the same region in the intact protein, whereas the data from *Cp*OGA crystal structure resulted in an RMSD value of 2.2 Å. Other OGA residues predicted by MD to make hydrogen bonds to the p53′ peptide include H433, D243, Y137 and K166; however, these were again short lived.

TAB1′: As seen in the case of poly-Ala′, in TAB1′ the most stable intermolecular hydrogen bond (24.3% occupancy) was again a peptide backbone interaction, specifically between A397(+2) and the sidechain of OGA residue D243. No other interactions existed for >8% of the total simulation time.

Structure and dynamics of the bound glycopeptides

To assess the presentation and conformation of each of the glycopeptides, obtained from the MD simulations, with that seen in the crystal structure of the p53 glycopeptide–*Cp*OGA complex (Schimpl et al. 2012), the glycan atoms in the glycopeptides were aligned onto the crystallographic coordinates of GlcNAc in the complex (Figure 1C). This alignment indicated that, in all cases, despite the difference in sequences, each peptide backbone adopted a similar "V"-shaped conformation; comparable with that seen for glycopeptide fragments of p53, TAB1 and hOGA in complex with *Cp*OGA (Schimpl et al. 2012). This "V"-shaped conformation is likely stabilized by a combination of intermolecular interactions, between the peptides and those *Bt*OGA residues identified from both the

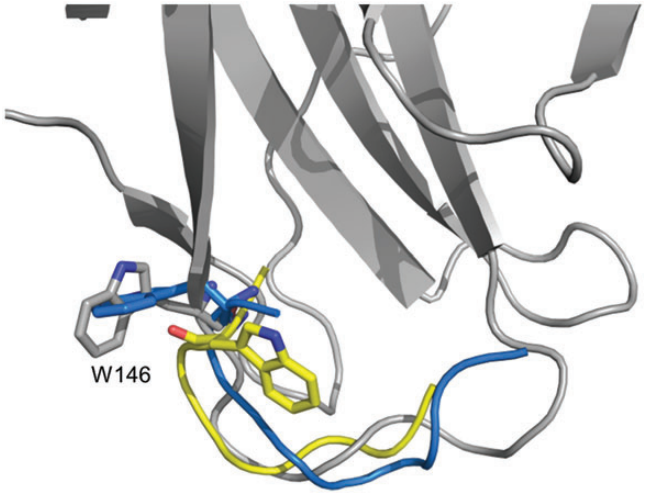

Fig. 2. Ribbon representation of both, the most populated p53′ glycopeptide (blue) and the p53 glycopeptide from the *Cp*OGA complex 2YDR (yellow), were aligned onto the corresponding sequence (Leu 145 to Gly 151) in the p53 protein crystal (PDBid 1TUP) (grey). Trp 146 of p53 (-3 position in the glycopeptides) is shown in stick with the GlcNAc residue removed for clarity.

hydrogen bond and the binding interaction analyses, as well as intramolecular bonds within the glycopeptide. In addition to the intermolecular interaction, discussed previously, an analysis of intramolecular interactions showed that all glycopeptides, with the exception of poly-Ala′, form, primarily, but not exclusively, mainchain hydrogen bonds between residues at the −1 or −2 positions with opposing residues at the +1, +3 or +4 positions in the glycopeptide (see Supplementary data, Table SIII). Specifically, in the complex of *Bt*OGA with the p53′-glycopeptide, a hydrogen bond was observed (3.0 Å, 76% occupancy) between the carboxylate of D(−1) and the mainchain nitrogen atom of T(+1). A similar interaction has been reported in the crystal structure of the p53 glycopeptide with *Cp*OGA (2.5 Å) (Schimpl et al. 2012). That this interaction formed spontaneously in the MD data, which was initiated independent of the *Cp*OGA–p53 complex, adds to the credibility of the general approach taken here to model the *Bt*OGA–glycopeptide complexes. Similarly, a bond was observed (2.9 Å, 53% occupancy) in the TAB1′–glycopeptide complex between the mainchain carbonyl oxygen atom in Y(−1) and the sidechain hydroxyl in S(+1) of the glycopeptide. Interestingly, this interaction is not seen in the crystal structure of the TAB1–glycopeptide in complex with *Cp*OGA (average distance 6.7 Å) (Schimpl et al. 2012); most likely due to the glycopeptide terminating at Ser at the +1 position, as opposed to being part of a longer peptide (10 residues) in our model. In the Histone H3′ glycopeptide, a weak hydrogen bond was seen (3.0 Å, 20% occupancy) between the −1 position (K, mainchain N) and the −4 position (T, mainchain carbonyl). Similarly, in the c-Myc′ glycopeptide, a weak hydrogen bond was predicted to exist (3.1 Å, 26% occupancy) between the backbone atoms at the −2 position (L, mainchain carbonyl) and the +3 position (L, mainchain N). In contrast to the cases of p53′ and TAB1′, in which the hydrogen bond spans only the glycosylated residue, in Histone H3′ and c-Myc′, the hydrogen bond spans three residues, presumably weakening the interaction in the latter cases.

With regard to intramolecular glycan–peptide interactions, it is notable that in each case a hydrogen bond was formed between the hydroxymethyl group (GlcNAc-O6) and the backbone of the glycosylated serine (Ser-N). This hydrogen bond (3.4 Å) was also seen in the p53′ glycopeptide–*Cp*OGA complex (Schimpl et al. 2012). GlcNAcylation via threonine, as in c-Myc′, appears to abolish this interaction.

An analysis of the RMSD of the peptide backbones was performed to quantify the conformational distribution for each peptide (Figure 3A), relative to the initial structure in the MD

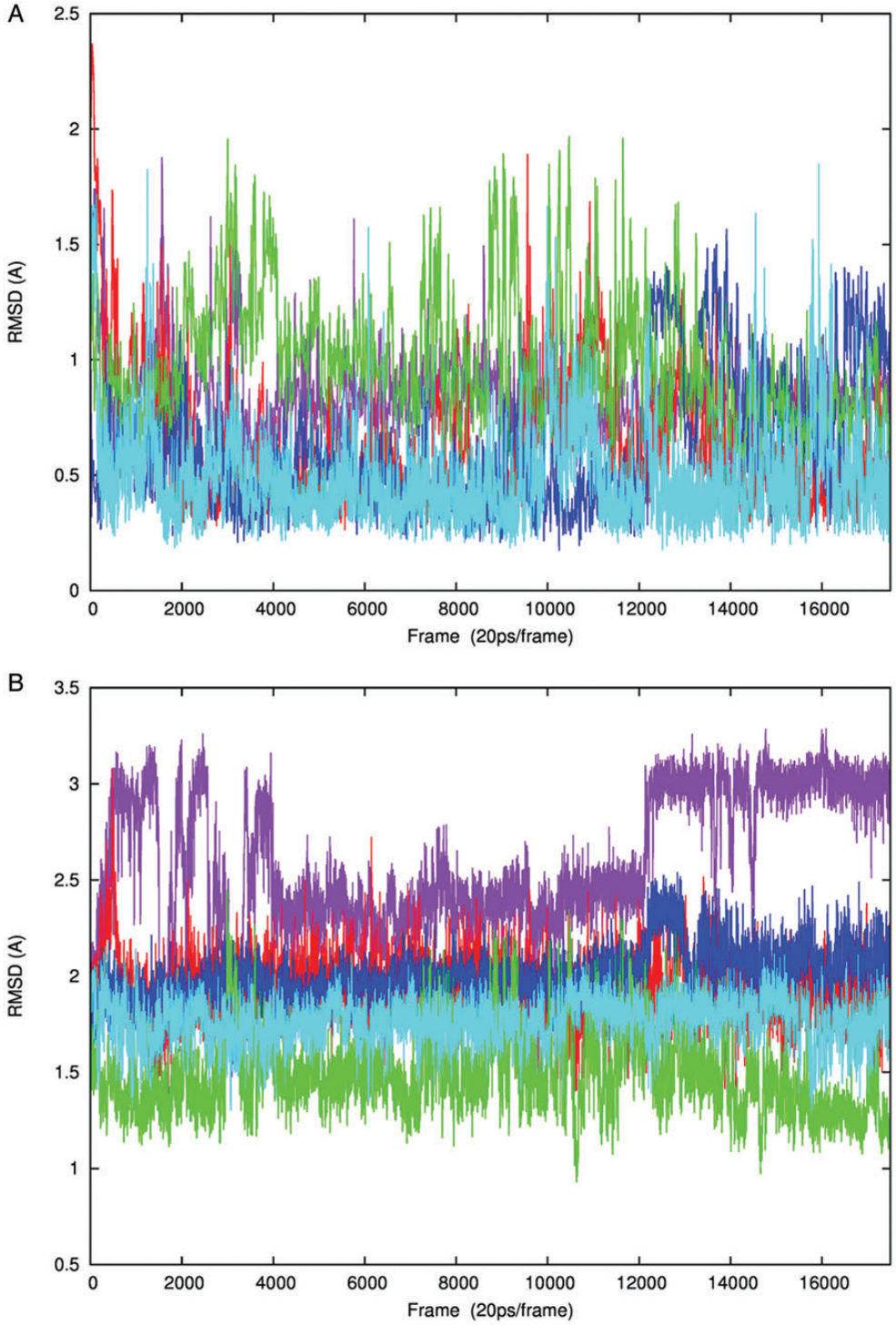

Fig. 3. Backbone atom (CA, C, O, N) RMSD values for each peptide throughout the 350ns MD as a function of time. **(A)** Average RMSD of each peptide relative to its starting conformation. **(B)** Each peptide relative to the backbone conformation of the p53 glycopeptide conformation from the crystal structure (PDBid 2YDR). The RMSD values are coloured as follows; c-Myc′ (purple), Histone H3′ (red), p53′ (blue), poly-Ala′ (green), TAB1′ (cyan).

simulations. This analysis indicated that Poly-Ala′ was the most flexible glycopeptide, with the average RMSD values being: 1.0 Å (poly-Ala′), 0.8 Å (c-Myc′), 0.6 Å (Histone H3′), 0.6 Å (p53′) and 0.5 Å (TAB1′). The RMSD analysis (Figure 3B), relative to the conformation of the p53–glycopeptide in complex with *Cp*OGA (Schimpl et al. 2012), indicated that the backbone conformations of each of the glycopeptides are similar: 2.6 Å (c-Myc′), 1.9 Å (Histone H3′), 2.0 Å (p53′), 1.5 Å (poly-Ala′) and 1.8 Å (TAB1′). Additionally, to identify similar conformational features sampled during the simulation of each glycopeptide, a clustering analysis was performed on the five MD data sets. The cluster analysis revealed that the top three clusters from each peptide cover over 75% of the total conformational space sampled. Of all of the glycopeptides, poly-Ala′ and c-Myc′ displayed the highest degree of structural diversity significantly populating at least two clusters (see Table III), whereas all of the other glycopeptides could be characterized by a single dominant cluster. The RMSD of each peptide, from the most populated cluster, relative to the conformation of the p53–glycopeptide in complex with *Cp*OGA (Schimpl et al. 2012), was comparable with those seen in the average RMSD analyses, and also showed that the peptide backbone conformations are all within 1.7 Å of each other (Table III).

Structure and dynamics of the free glycopeptides

To gain an understanding of the extent to which the free glycopeptides in solution exist in conformations appropriate for binding to the *Bt*OGA, independent MD simulations (500 ns) were performed on each of the five glycopeptides. An analysis of the intramolecular hydrogen bonds in each system indicated that p53′ and TAB1′ displayed transient interactions for ∼29% of the simulations (see Supplementary data, Table SIV). The other glycopeptides appeared to be even less-ordered. The interactions within p53′ were similar to those seen in the bound peptide, whereas in TAB1′ they were unique to the free peptide. Thus, only in the case of p53′ does the free peptide appear to populate to some extent conformations comparable with those seen in the bound state. It is notable also that intramolecular interactions between the glycan and residues at the +1 or −1 positions were observed in the free glycopeptides, similar to those seen in the bound forms, but that even in the case of p53′ these interactions were transient.

To further quantify the conformational similarities between the bound and free glycopeptides, the bound conformation of each peptide, as extracted from the dominant cluster, was compared with the solution conformations. The resulting RMSD plots for the free glycopeptides, relative to the bound conformaiton, indicate that p53′ exhibits relatively low RMSD values (of ∼2–4 Å), further supporting the idea that the p53′ glycopeptide exists in a conformation similar to that of the bound (Supplementry data, Figure S3). In contrast, in the case of TAB1′, the internal hydrogen bonds for the free glycopeptide are associated primarily with a conformation that is significantly different from that seen in the bound state (RMSD ∼4–5 Å). Neither histone H3′ nor poly-Ala′ appear to adopt conformations similar to the bound forms, however, c-Myc′, despite the absence of stable internal hydrogen bonds, also adopts conformations in solution that are similar to the bound form (RMSD ∼2–5 Å).

Table III. Clustering analysis[a] for each of the five glycopeptides

Cluster	Cluster population	RMSD
c-Myc		
1	56	2.9
2	30	2.3
3	6	3.0
Total	91	
Histone H3		
1	79	1.9
2	9	1.7
3	5	2.0
Total	92	
p53′		
1	92	1.9
2	5	2.2
3	2	2.2
Total	99	
Poly-Ala′		
1	36	1.4
2	27	1.4
3	12	1.3
Total	75	
TAB1′		
1	91	1.8
2	5	1.5
3	2	1.5
Total	98	

[a]Relative to the backbone conformation of the p53 glycopeptide in complex with *Cp*OGA (Schimpl et al. 2012) and the structure with the median RMSD in the cluster. Populations in percent, distances in Angstroms.

Binding free energy analysis

Per-residue binding free energy calculations (molecular mechanics–generalized born surface area (MM–GBSA)) were carried out on GlcNAc-OMe in complex with *Bt*OGA, as well as the five glycopeptide–*Bt*OGA complexes. This allowed us to separate and quantify the contributions made by the GlcNAc and the peptide portions of the glycopeptides, as well as to identify which *Bt*OGA residues make a significant contribution to glycopeptide affinity.

Bt*OGA per-residue binding energy contributions*. In agreement with the hydrogen bond analysis of the OGA–glycopeptide interactions, the MM–GBSA data showed that the five *Bt*OGA residues (D344, N339, D242, K166 and H433), which make sidechain interactions with the glycan, rank among the highest contributors to binding affinity (Figure 4 and Supplementary data, Table SV). Additionally, the MM–GBSA analysis for the OGA–GlcNAc-OMe complex found that protein residues (W337, Y282, N372, Y137, V314 and V342) contribute toward binding GlcNAc through a combination of van der Waals and hydrophobic interactions. Other residues that were predominantly associated with the peptide portion of the glycopeptides were also identified; for example D243 (see Figure 4 and Supplementary data, Table SV). Notably, residue Y137 showed a markedly greater contribution to binding for all glycopeptides, compared with that seen for its interaction with GlcNAc-OMe alone. Further, the contribution from Y137 varied among the glycopeptides, making a significantly stronger interaction with p53′ (−8.6 kcal/mol), which may be compared with values for histone H3′ (−5.5 kcal/mol), c-Myc′ (−5.0 kcal/mol), TAB1′ (−4.5 kcal/mol), poly-Ala′ (−3.9 kcal/mol) and GlcNAc-OMe (−1.2 kcal/mol).

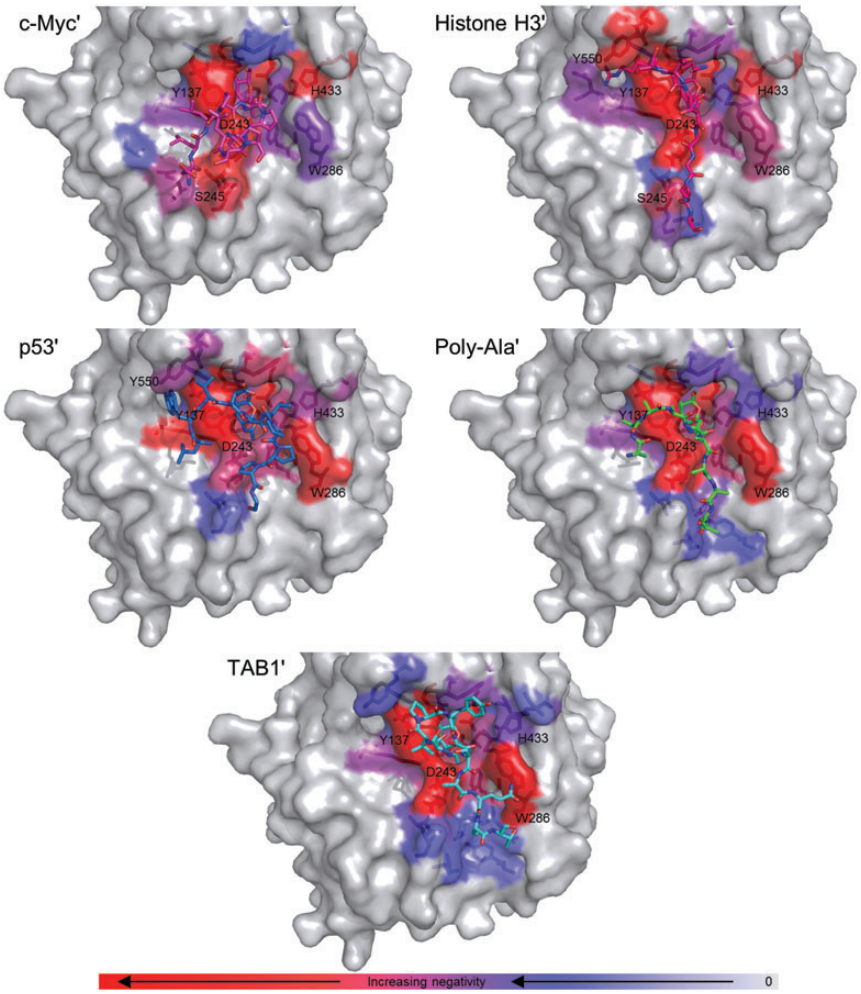

Fig. 4. The middle structure from each cluster analysis, of c-Myc′ (purple), Histone H3′ (red), p53′ (blue), poly-Ala′ (green) and TAB1′ (cyan) are shown in complex with OGA. The per-residue binding interactions for the OGA residues are represented as a colour scale; from lowest (most attractive) (red) to highest (least attractive) (grey) binding interaction energy value. The OGA residues that are most significant to peptide binding are shown for each complex.

This enhancement stems from hydrogen bonding with the glycosylation site Ser or Thr residues, common to each glycopeptide, as well as through van der Waals interactions with glycopeptide residues at the −2 and −3 positions. Similarly, residue W286 contributes to peptide binding in all cases, principally through van der Waals interactions with residues at the +2 to +4 positions (TAB1′ (−5.0 kcal/mol) > poly-Ala′ (−3.7 kcal/mol) > p53′ (−3.1 kcal/mol) histone H3′ (−2.0 kcal/mol) > c-Myc′ (−1.3 kcal/mol) > GlcNAc-OMe (−0.2 kcal/mol)). Residue H433 also contributes to binding in all of the peptides; interestingly the interactions with H433 ranks the glycopeptides in an inverse order relative to that seen for W286, namely c-Myc′ (−2.6 kcal/mol) > histone H3′ (−2.5 kcal/mol) > p53′ (−1.7 kcal/mol) > TAB1′ (−1.3 kcal/mol) > poly-Ala′ (−1.0 kcal/mol) > GlcNAc-OMe (−0.7 kcal/mol). This suggests that W286 and H433 counter balance each other in terms of peptide specificity.

Certain *Bt*OGA residues were also found to be of particular importance to a subset of glycopeptides (Figure 4). For example, residues Y550 (−2.9 kcal/mol), R347 (−2.3 kcal/mol) and S245 (−2.3 kcal/mol) each contribute significantly to the affinity of histone H3′. Residue Y550 interacts favorably with histone H3′ residue R(−2), primarily through van der Waals interactions. As might be expected, the interaction between R347 and R(−2) is electrostatically repulsive; however, it is predicted to be net favorable, as a result of desolvation effects. This suggests that mutation of R347, to either D or E, or to an aromatic residue, may lead to increased binding specificity for histone H3′. As in the case of histone H3′, residue Y550 also contributes to the affinity of p53′ (−1.7 kcal/mol), primarily through van der Waals interactions with W(−3). Finally, residues S245 (−2.4 kcal/mol) > I244 (−2.7 kcal/mol) > Q210 (−1.8 kcal/mol) are each involved in favorable interactions with c-Myc′ at E(−4).

Glycopeptide per-residue binding energy contributions. In order to define the contributions from each residue in the five glycopeptides, a per-residue decomposition of the MM–GBSA data was performed (see Table IV for total energy interactions and Supplementary data, Table SVI for the per-residue breakdown). Specifically, over 50% of the total binding free energy for each of the five glycopeptides comes directly from interactions involving

Table IV. Total component binding free energies[a] for each of the five glycopeptides.

Moiety	Van der Waals	Electrostatic	Polar Desolvation	Nonpolar Desolvation	Total Interaction
c-Myc′					
GlcNAc	−13.0	−45.9	34.3	−2.4	−27.0 (2.2)
Peptide	−21.8	37.9	−32.5	−2.9	−19.1 (6.0)
Histone H3′					
GlcNAc	−12.9	−46.8	34.2	−2.3	−27.9 (1.7)
Peptide	−25.9	−175.6	180.7	−3.8	−24.4 (7.5)
p53′					
GlcNAc	−13.1	−45.6	33.6	−2.3	−27.4 (1.7)
Peptide	−31.0	51.2	−39.8	−4.6	−24.2 (7.0)
Poly-Ala'					
GlcNAc	−12.3	−41.5	33.2	−2.3	−22.9 (4.7)
Peptide	−23.8	−3.8	12.2	−3.5	−18.7 (7.6)
TAB1′					
GlcNAc	−13.5	−40.4	32.4	−2.3	−23.8 (1.9)
Peptide	−27.0	−13.6	20.4	−4.2	−24.4 (7.9)

[a]In kcal/mol with standard deviation in parentheses

O-GlcNAc. The significance of the GlcNAc moiety in driving glycopeptide binding has similarly been shown through kinetic studies of hOGA with various O-GlcNAcylated substrates (Shen et al. 2012). Interestingly, the binding energy from the GlcNAc falls into two distinct categories. For the glycopeptides; c-Myc′, Histone H3′ and p53′, the GlcNAc contributed essentially equivalent interaction energies (−27.0 to −27.9 kcal/mol), whereas for the poly-Ala' and TAB1′ glycopeptides, the glycan contributed −22.9 and −23.8 kcal/mol, respectively (Table IV). The loss in affinity in the latter two cases is a direct result of weakened electrostatic interactions (decreased by 4–5 kcal/mol) with the glycan. The analysis of the hydrogen bond interactions for these systems suggests that this is a result of weaker interactions specifically with residues D344 and N339, as indicated by both slightly lower occupancy and larger distances of the hydrogen bonds (Supplementary data, Table SII). Closer inspection indicates that the glycan, in the poly-Ala' and TAB1′ glycopeptide complexes, is more mobile than in the other three complexes; the average positional RMSD for the nonhydrogen atoms in GlcNAc for the poly-Ala' (0.82 Å) and TAB1′ (0.61 Å) complexes are slightly higher than for the GlcNAc in the other three complexes (c-Myc′: 0.53 Å, histone H3′: 0.41 Å, and p53′: 0.51 Å).

The peptides (not including the GlcNAc) contribute (−18.7 to −24.4 kcal/mol) to the total interaction energy; or up to ∼50% of the total binding affinity. As might be expected on the basis of relative flexibilities, the interaction energies associated with the GlcNAc residues are much less variable than those of the peptides. These data suggest that peptide residues in the vicinity of the glycosylation site (±5 residues) can enhance glycopeptide affinity, and also that loss of the glycan from the peptide, by enzymatic cleavage, would markedly reduce the affinity of the products. Notably, as in the case of the GlcNAc residues, the glycosylation site contributed similar values in each complex (an average of −2.9 ± 0.7 kcal/mol), with no significant differences being observed for Ser or Thr (c-Myc′) residues. In the poly-Ala' reference system, the residues in the +1 to +5 positions, relative to the glycosylation site, contribute ∼−10.0 kcal/mol to the binding interaction. In the cases of the other four glycopeptides, residues in the +1 to +5 positions make weaker interactions with *Bt*OGA (histone H3′: −8.0 kcal/mol, p53′: −7.8 kcal/mol, c-Myc′: −4.9 kcal/mol), relative to poly-Ala', with the exception of TAB1′ (TAB1′: −11.1 kcal/mol). In contrast, relative to poly-Ala' (−6.2 kcal/mol), residues in the −1 to −4 positions consistently contribute significantly to the interaction energies (p53′: −16.0 kcal/mol, histone H3′: −13.0 kcal/mol, c-Myc′: −11.5 kcal/mol, TAB1′: −10.3 kcal/mol).

Discussion

Here we have extended the simulations to systems for which experimental data do not exist. The new structural information provides a fundamental understanding of the manner in which OGA tolerates variations in peptide substrate sequence. It is also important to note that interaction energies are highly sensitive to the 3D structure of the complex, and for these reasons we have placed considerable emphasis on validating and interpreting the structural properties of the *Bt*OGA–glycopeptide complexes.

The computational analyses indicate that each of the five glycopeptides can adopt comparable conformations when bound to the enzyme and that the interaction energies of each complex were remarkabley similar, spanning a range of <5 kcal/mol. The bound complexes are in good agreement with recently reported data for the complex between the bacterial OGA homolog from *C. perfringens* (*Cp*OGA) in complex with an O-GlcNAcylated p53 peptide (PDBid 2YDR) (Schimpl et al. 2012). The fact that poly-Ala' behaved similarly to the other four natural glycopeptides suggests that the interactions with *Bt*OGA are structurally nonspecific. The occupancy analysis of the hydrogen bond network revealed that none of the peptides formed stable inter- or intramolecular hydrogen bonds. This is not unexpected, given the highly dynamic nature of the peptides. Nevertheless, the analyses suggested that the "V"-shaped conformation of each of the glycopeptides was stabilized to greater or lesser extents by weak intramolecular interactions spanning the glycosylation site, as well as by a limited number of intermolecular interactions.

Although stacking between Y137 and the peptide backbone was noted by Shimpl et al., no such stacking was observed during the simulations of the *Bt*OGA glycopeptide complexes (Supplementary data, Table SVII). Rather, the interaction energy associated with Y137 stems from less-ordered vdW contacts with the glycopeptides. The lack of stacking during the MD simulations indicates that there is no hydrophobic or van der Waals driving force for such an orientation. The lack of a stable stacked geometry might also arise from the inability of the classical force field to capture effects arising from π to π orbital overlap.

A per-residue analysis of the energetic contributions of *Bt*OGA to binding, from both hydrogen bonding and nonpolar interactions, indicated that a small subset of residues (D344, N339, D242, K166, H433, W337, Y282, N372, Y137, V314 and V342) is responsible for binding the glycan. Similarly, the analysis indicated that four *Bt*OGA residues (Y137 > D243 ≈ W286 > H433) were predominantly responsible for binding the peptide portion for all of the glycopeptides. Our results are in good agreement with the experimentally observed loss of enzymatic activity observed for *Bt*OGA mutants (D242A/N, D243A/N, N372A,

Y137F and Y282F) (Dennis et al. 2006), as well as *Cp*OGA mutant D401A (*Bt*OGA D344) (Rao et al. 2006). It has also been reported that mutation of the conserved Y69 in hOGA (*Bt*OGA Y137) to any residue other than F led to a loss in enzymatic activity (Schimpl et al. 2012). Given that the remaining peptide residues contribute to affinity through both backbone and sidechain interactions, it is unclear to what extent mutations at these positions would affect affinity or modulate enzyme K_m values. The per-residue analysis also identified residues which contributed specifically to affinity for each of the glycopeptides. Such data enable the prediction of sites whose mutation may lead to significant preferences for one substrate over another. For example, mutation of *Bt*OGA R347 to either a negatively charged, or aromatic, residue could potentially lead to an increased affinity and specificity for histone H3′.

Taken together, the MM–GBSA data indicate that the affinity for substrate varies as a function of peptide sequence, despite the fact that there are no strong recognition motifs associated with any of the peptides. This observation would be expected to lead to variations in the enzyme K_m values, as observed experimentally (Schimpl et al. 2012). Consistent with interpretations of the crystal structures (Schimpl et al. 2012), affinity contributions are made by peptide residues near the GlcNAc site, but here we also see significant contributions from residues as distal as the −4 position. Although the glycopeptides all adopted grossly similar "V"-shaped conformations, none displayed persistent internal hydrogen bonds. Thus it is uncertain to what extent internal structuring of the peptide contributes to the affinity of the interactions, or to the K_m of the enzyme, as has been proposed (Schimpl et al. 2012).

Conclusion

Here we have presented an atomic-level analysis of the structure and dynamics of the binding interaction between a bacterial OGA homolog in complex with five GlcNAcylated peptides of biological interest. The primary motivation for the research was to develop an understanding of the unique lack of substrate sequence specificity displayed by OGA. The MD simulations show that despite the differences in sequence, all of the glycopeptides adopted a similar conformation when bound to OGA, consistent with limited experimental data (Schimpl et al. 2012), but nevertheless retain considerable internal flexibility. This flexibility is also thought to occur in the region spanning the sites of known O-GlcNAcylated proteins, and as such there are currently no protein structures available, with the exception of p53 (PDbid 1TUP), where these regions are fully defined (Schimpl et al. 2012). This observation has implications not only for the action of the hydrolase OGA, but also potentially for the specificity of its cognate partner, the transferase OGT.

The theoretical approach taken here enabled us to quantify the relative importance of specific and nonspecific interactions. The data show that the specificity of OGA for the GlcNAcylated peptides arises primarily from strong conserved interactions with the glycan; nevertheless, nonspecific interactions with the peptidic component of the substrates contribute significantly to affinity. This observation provides a clear explanation for the ability of this enzyme to tolerate a range of polymorphisms in the vicinity of the glycosylation site in the substrate. The analysis also indicates that variations in observed K_m values (Schimpl et al. 2012) are a natural consequence of affinity contributions from nonspecific interactions with the substrate. The potential exists therefore to employ the present analysis to guide the development glycopeptide-specific inhibitors, or conversely, the conversion of OGA into a reagent that could target specific O-GlcNAcylated peptide sequences.

Computational methods

MD simulations

All MD simulations were carried out using version 4.04 of the Groningen machine for chemical simulations (GROMACS) (Hess et al. 2008) or assisted model building with energy refinement (AMBER)12 (Case et al. 2005) software packages. All calculations employed the AMBER99SB (Cornell et al. 1995) parameters for the protein augmented by the glycans and glycoproteins in AMBER (GLYCAM)06g (Kirschner et al. 2008) parameters for the carbohydrate. The total charge was neutralized by addition of explicit Na^+ or Cl^- counter ions as required. Each system was solvated with TIP3P water (Jorgensen et al. 1983) in a cubic simulation box (120 Å per side). The positions of all water molecules, counterions and all hydrogen atoms were optimized through 10,000 steps of steepest descent energy minimization. During energy minimization and MD equilibration the nonhydrogen atoms were restrained with a force constant of 1000 kJ/mol/nm^2.

A step-wise protocol was employed for MD equilibration beginning with heating in the NVT ensemble to 300 K over 100 ps. Subsequently, water, hydrogen atoms and counter ions were equilibrated for a further 200 ps in the NPT ensemble at 300 K at a pressure of 1 bar. Restraints were then removed from the sidechains of the protein–ligand complexes and the system equilibrated for a further 1 ns. During the final equilibration phase (2 ns), all atoms were released from positional restraints. In all simulations, the linear constraint solver (LINCS) was used to constrain all hydrogen-containing molecules (Hess et al. 1997), permitting an integration step of 2 fs. The temperature was held constant at 300 K by a Langevin thermostat with a coupling time constant of 0.1 ps. Electrostatic interactions were treated with the Particle–Mesh Ewald (PME) algorithm (Darden et al. 1993). Cut-off values for Coulomb interactions were set to 12 Å, while van der Waals interactions were switched off between 10 and 11 Å.

Protein coordinates were obtained from the crystal structure of the bacterial OGA from *B. thetaiotaomicron* (*Bt*OGA), in complex with the GlcNAc transition state mimic inhibitor PUGNAc (PDBid: 2VVS) (Macauley et al. 2008). The N-terminal domain between L5 and P125, located between 28 Å and 54 Å from the binding site, was removed to reduce the computational demand. The N- and C-terminal residues (Ser126 and Ala586) were capped with acetyl (ACE) and *N*-methyl (NME) groups, respectively. Aside from residue D243 (D175 in hOGA), which is known to be protonated (Çetinbaş et al. 2006), the protonation state of all other ionizable residues was chosen to reflect their standard protonation at pH 7. All crystallographic water molecules were removed and missing hydrogen atoms were added to the structure using pdb2gmx (Feenstra et al. 1999). Where possible, protonation of His residues at the - or δ-nitrogen atoms was based on optimal hydrogen bonding interactions; otherwise a default of His-δ was assumed.

Generation of the glycopeptide–BtOGA complexes

The conformation of the bound native GlcNAc ligand was unambiguously determined based on the PUGNAc coordinates. It has previously been shown that the common *anti*-conformation of the NAc group in PUGNAc does not permit it to fit into the binding site of the *Bt*OGA (Rao et al. 2006), thus the NAc group of the GlcNAc residue was rotated from *anti* to *syn*, comparable to the orientation seen for the oxime in the *Bt*OGA–PUGNAc complex (Figure 5A). *Syn* and *anti* refer to values of 0° and 180° with respect to the H2-C2-C-N torsion angle; however, as hydrogen atoms are not routinely reported in crystal structures, these orientations have alternatively been defined here in terms of C1-C2-N-C for which *syn* = −60° and *anti* = 120°. A 3D structure for GlcNAc-β-O-Ser was built using the GLYCAM webtool (Woods 2005–2013) and modified using Maestro (Maestro 2011) to generate three rotamers of the serine side chain ($\chi_1 = 40.7$, 170.6, and −49.1°, see Figure 5B), as identified from the Dynameomics database (Day et al. 2003; Beck et al. 2008; Kehl et al. 2008; Simms et al. 2008). The N- and C-terminal residue was capped for all glycopeptides with an acetyl (ACE) or *N*-methyl (NME) group, respectively. To determine the preferred conformation of the GlcNAc-β-O-Ser linkage, three initial MD simulations of this glycopeptide in complex with *Bt*OGA were carried out. Each simulation was initiated with the χ_1 angle of the Ser sidechain in one of the three possible stable rotamers.

Following MD simulations, the most populated GlcNAc-β-O-Ser rotamer was identified ($\phi/\psi/\chi_1 \approx -61.3/-179.5/-77.9°$; see Figure 5C and Supplementary data, Figure 1), and used to define an initial orientation of the GlcNAc moiety relative to the peptide backbone for use in docking each of the GlcNAcylated peptides (c-Myc′, histone H3′, p53′, TAB1′ and a poly-Ala′) to *Bt*OGA.

Molecular docking was performed with Autodock Vina (AD Vina) (Trott and Olson 2010), employing the iterated local search global optimizer algorithm (Baxter 1981; Abagyan et al. 1994). Structures for the five fully extended glycopeptides were generated using Maestro (Maestro 2011), and docked into the *Bt*OGA structure, while maintaining the GlcNAc-β-O-Ser linkage at the preferred rotamer values, as defined above. All other backbone and sidechain torsion angles in the glycopeptides were allowed to be flexible. The graphical user interface program AutoDock Tools (Sanner 1999) was used to prepare the structures for all docking simulations. To prepare the *Bt*OGA crystal structure for docking, crystallographic water molecules, as well as the co-crystallized inhibitor, PUGNAc, were removed. Polar hydrogen atoms were added where necessary with AutoDock Tools. The docking region was selected by centering a grid box (35 Å per side) on the Cα atom of the nucleophile residue D243. The optimal glycopeptide conformation from each of the five dockings was determined by searching the ranked docking poses for the lowest energy structure; in which the GlcNAc was placed in a pose comparable with that seen in the MD simulations of the GlcNAc-β-O-Ser complex (Figure 1B). The five *Bt*OGA–glycopeptide complexes were employed for subsequent MD simulations. Simulations of the free glycopeptides were initiated from the most populated conformation of the bound glycopeptides.

Analysis

The conformational properties of each of the five peptides was analyzed in terms of RMSD of the peptide backbone atoms (Cα, N, C, O) employing visual molecular dynamics (VMD) (Humphrey et al. 1996). The RMSD values for each peptide was calculated relative to the first time step of the MD simulation, and also with reference to the experimental coordinates reported recently for the p53 glycopeptide bound to *Cp*OGA. The Cremer–Pople-puckering parameters (Cremer and Pople 1975) of the GlcNAc ring were monitored throughout each MD using the GROMACS (Hess et al. 2008) tool g_puckering. Clustering analyses of the peptide conformations were performed using the GROMACS tool g_cluster, employing the Gromos Algorithm (Daura et al. 1999) using an RMSD cut-off of 1.0 Å. Stacking interactions were monitored using the GROMACS (Hess et al. 2008) tool g_sgangle to measure the angle and distance between two specified groups. An analysis of the hydrogen bond network was carried out for each complex using VMD (Humphrey et al. 1996) with a distance cut-off of 3.5 Å and a minimal occupancy of 5%. Binding free energies were calculated with the single-trajectory MM–GBSA method (Kollman et al. 2000, Srinivasan et al. 1998), with average values computed from an ensemble of 100 uncorrelated snapshots collected every 3.5 ns from the 350 ns trajectory. Before the analyses, all water molecules were removed from each complex, and the solvation energy approximated through the generalized Born (GB) implicit solvation model (igb = 2) (Onufriev et al. 2004). This approach has been applied by us (Kadirvelraj et al. 2006, Kadirvelraj et al. 2011) and others (Bryce et al. 2001) for the analyses of carbohydrate–protein affinity.

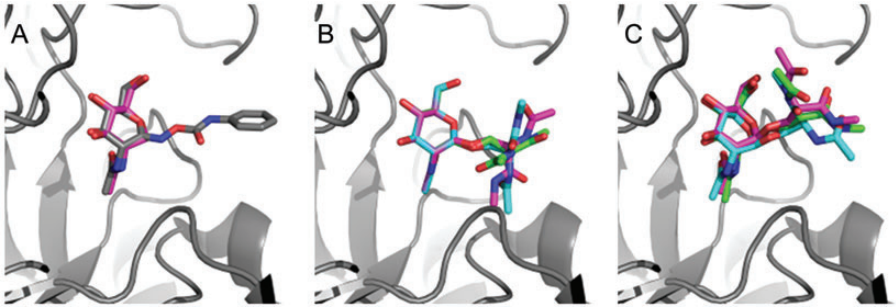

Fig. 5. (A) β-OMe-GlcNAc(z), shown as sticks with pink carbons, was superimposed onto PUGNAc, shown as grey sticks, in the OGA (PDBid 2VVS) binding site, shown as grey cartoon. **(B)** The three β-Ser-O-GlcNAc(z) g+ (green), t (cyan), g- (pink) rotamers, prior to MD. **C)** β-Ser-O-GlcNAc(z) rotamers after 10 ns simulation.

Supplementary Data

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Funding

This work was supported by the Science Foundation of Ireland (08/IN.1/B2070) and the European Union (ERDF) for financial support, as well as the Irish Research Council for Science, Engineering and Technology, together with Hewlett & Packard (RCS932).

Acknowledgements

We would like to thank Prof. Daan van Aalten, University of Dundee, Scotland for providing us with the crystal structure 2YDR prior to its deposition in the PDB. Support is also acknowledged from the National Institutes for Health (R01 GM094919 (EUREKA), P41 GM103390).

Conflict of interest

None declared.

Abbreviations

ACE, acetyl; AMBER, assisted model building with energy refinement; *Bt*OGA, *Bacteroides thetaiotaomicron* GH84; *Cp*OGA, *Clostridium perfringens*; GB, generalized Born; GLYCAM, glycans and glycoproteins in AMBER; hOGA, human OGA; LINCS, linear constraint solver; MD, molecular dynamics; MM–GBSA, molecular mechanics–generalized born surface area; NME, *N*-methyl; O-GlcNAc, O-linked *N*-acetylglucosamine; O-GlcNAcase or OGA, O-GlcNAc hydrolase; OGT, O-GlcNAc transferase; PDB, Protein Data Bank; PME, Particle–Mesh Ewald; GROMACS, Groningen machine for chemical simulations; VMD, visual molecular dynamics; OMe, O-methyl.

References

Abagyan R, Totrov M, Kuznetsov D. 1994. ICM—A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem*. 15:488–506.

Baxter J. 1981. Local optima avoidance in depot location. *J Oper Res Soc*. 32:815–819.

Beck DA, Jonsson AL, Schaeffer RD, Scott KA, Day R, Toofanny RD, Alonso DO, Daggett V. 2008. Dynameomics: mass annotation of protein dynamics and unfolding in water by high-throughput atomistic molecular dynamics simulations. *Protein Eng Des Sel*. 21:353–368.

Bryce RA, Hillier IH, Naismith JH. 2001. Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys J*. 81:1373–1388.

Case DA, Cheatham III TE, Darden TA, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling CL, Wang B, Woods RJ. 2005. The AMBER biomolecular simulation programs. *J Comput Chem*. 26:1668–1688.

Çetinbaş N, Macauley MS, Stubbs KA, Drapala R, Vocadlo DJ. 2006. Identification of Asp174 and Asp175 as the key catalytic residues of human O-GlcNAcase by functional analysis of site-directed mutants. *Biochemistry (Mosc)*. 45:3835–3844.

Cheng X, Cole RN, Zaia J, Hart GW. 2000. Alternative O-glycosylation/O-phosphorylation of the murine estrogen receptor β. *Biochemistry (Mosc)*. 39:11609–11620.

Cho Y, Gorina S, Jeffrey PD, Pavletich NP. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science*. 265:346–355.

Chou T-Y, Hart GW, Dang CV. 1995. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J Biol Chem*. 270:18961–18965.

Comer FI, Hart GW. 2000. O-Glycosylation of nuclear and cytosolic proteins: Dynamic interplay between O-GlcNAc and O-phosphate. *J Biol Chem*. 275:29179–29182.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*. 117:5179–5197.

Cremer D, Pople JA. 1975. General definition of ring puckering coordinates. *J Am Chem Soc*. 97:1354–1358.

Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N [center-dot] log(N) method for Ewald sums in large systems. *J Chem Phys*. 98:10089–10092.

Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. 1999. Peptide folding: When simulation meets experiment. *Angew Chem Int Ed*. 38:236–240.

Day R, Beck DA, Armen RS, Daggett V. 2003. A consensus view of fold space: combining SCOP, CATH, and the Dali Domain Dictionary. *Protein Sci*. 12:2150–2160.

Dennis RJ, Taylor EJ, Macauley MS, Stubbs KA, Turkenburg JP, Hart SJ, Black GN, Vocadlo DJ, Davies GJ. 2006. Structure and mechanism of a bacterial [beta]-glucosaminidase having O-GlcNAcase activity. *Nat Struct Mol Biol*. 13:365–371.

Feenstra KA, Hess B, Berendsen HJC. 1999. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem*. 20:786–798.

Hanover JA. 2001. Glycan-dependent signaling: O-linked N-acetylglucosamine. *FASEB J*. 15:1865–1876.

Hanover JA, Krause MW, Love DC. 2010. The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine. *Biochim Biophys*. 1800:80–95.

Hart GW, Housley MP, Slawson C. 2007. Cycling of O-linked [beta]-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature*. 446:1017–1022.

He Y, Martinez-Fleites C, Bubb A, Gloster TM, Davies GJ. 2009. Structural insight into the mechanism of streptozotocin inhibition of O-GlcNAcase. *Carbohydr Res*. 344:627–631.

Henrissat B, Davies G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol*. 7:637–644.

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. 1997. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*. 18:1463–1472.

Hess B, Kutzner C, van der Spoel D, Lindahl E. 2008. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 4:435–447.

Humphrey W, Dalke A, Schulten K. 1996. VMD: Visual molecular dynamics. *J Mol Graph*. 14:33–38.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*. 79:926–935.

Kadirvelraj R, Gonzalez-Outeriño J, Foley BL, Beckham ML, Jennings HJ, Foote S, Ford MG, Woods RJ. 2006. Understanding the bacterial polysaccharide antigenicity of *Streptococcus agalactiae* versus *Streptococcus pneumoniae*. *PNAS*. 103:8149–8154.

Kadirvelraj R, Grant OC, Goldstein IJ, Winter HC, Tateno H, Fadda E, Woods RJ. 2011. Structure and binding analysis of *Polyporus squamosus* lectin in complex with the Neu5Acα2–6Galβ1–4GlcNAc human-type influenza receptor. *Glycobiology*. 21:973–984.

Kamemura K, Hayes BK, Comer FI, Hart GW. 2002. Dynamic interplay between O-glycosylation andO-phosphorylation of nucleocytoplasmic proteins: Alternative glycosylation/phosphorylation of Thr-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J Biol Chem*. 277:19229–19235.

Kehl C, Simms AM, Toofanny RD, Daggett V. 2008. Dynameomics: a multi-dimensional analysis-optimized database for dynamic protein data. *Protein Eng Des Sel*. 21:379–386.

Kim EJ, Perreira M, Thomas CJ, Hanover JA. 2006. An O-GlcNAcase-specific inhibitor and substrate engineered by the extension of the N-acetyl moiety. *J Am Chem Soc*. 128:4234–4235.

Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J Comput Chem*. 29:622–655.

Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, et al. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc Chem Res*. 33:889–897.

Macauley MS, Bubb AK, Martinez-Fleites C, Davies GJ, Vocadlo DJ. 2008. Elevation of global O-GlcNAc levels in 3T3-L1 adipocytes by selective inhibition of O-GlcNAcase does not induce insulin resistance. *J Biol Chem*. 283:34687–34695.

Macauley MS, Vocadlo DJ. 2010. Increasing O-GlcNAc levels: An overview of small-molecule inhibitors of O-GlcNAcase. *Biochim Biophys Acta*. 1800:107–121.

Macauley MS, Whitworth GE, Debowski AW, Chin D, Vocadlo DJ. 2005. O-GlcNAcase uses substrate-assisted catalysis. *J Biol Chem*. 280:25313–25322.

Maestro. 2011. *Suite 2011: Maestro, version 9.2*, Schrödinger, LLC: New York, NY, 2011.

Onufriev A, Bashford D, Case DA. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*. 55:383–394.

Pathak S, Borodkin VS, Albarbarawi O, Campbell DG, Ibrahim A, van Aalten DM. 2012. O-GlcNAcylation of TAB1 modulates TAK1-mediated cytokine release. *EMBO J*. 31:1394–1404.

Perreira M, Kim EJ, Thomas CJ, Hanover JA. 2006. Inhibition of O-GlcNAcase by PUGNAc is dependent upon the oxime stereochemistry. *Bioorg Med Chem*. 14:837–846.

Rao FV, Dorfmueller HC, Villa F, Allwood M, Eggleston IM, van Aalten DMF. 2006. Structural insights into the mechanism and inhibition of eukaryotic O-GlcNAc hydrolysis. *EMBO J*. 25:1569–1578.

Sanner MF. 1999. Python: A programming language for software integration and development. *J Mol Graph Model*. 17:57–61.

Schimpl M, Borodkin VS, Gray LJ, van Aalten DMF. 2012. Synergy of peptide and sugar in O-GlcNAcase substrate recognition. *Chem Biol*. 19:173–178.

Shen DL, Gloster TM, Yuzwa SA, Vocadlo DJ. 2012. Insights into O-GlcNAc processing and dynamics through kinetic analysis of O-GlcNAc transferase and O-GlcNAcase activity on protein substrates. *J Biol Chem*. 287:15395–15408.

Simms AM, Toofanny RD, Kehl C, Benson NC, Daggett V. 2008. Dynameomics: design of a computational lab workflow and scientific data repository for protein simulations. *Protein Eng Des Sel*. 21:369–377.

Slawson C, Copeland RJ, Hart GW. 2010. O-GlcNAc signaling: a metabolic link between diabetes and cancer?. *Trends Biochem Sci*. 35:547–555.

Srinivasan J, Miller J, Kollman PA, Case DA. 1998. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J Biomol Struct Dyn*. 16:671–682.

Torres CR, Hart GW. 1984. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem*. 259:3308–3317.

Trott O, Olson AJ. 2010. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 31:455–461.

Whitworth GE, Macauley MS, Stubbs KA, Dennis RJ, Taylor EJ, Davies GJ, Greig IR, Vocadlo DJ. 2006. Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: Mechanistic and structural insights into inhibitor selectivity and transition state poise. *J Am Chem Soc*. 129:635–644.

Woods G. 2005–2013. GLYCAM Web.

Yang WH, Kim JE, Nam HW, Ju JW, Kim HS, Kim YS, Cho JW. 2006. Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nat Cell Biol*. 8:1074–1083.

Zeidan Q, Hart GW. 2010. The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. *J Cell Sci*. 123:13–22.

Zhang S, Roche K, Nasheuer H-P, Lowndes NF. 2011. Modification of histones by sugar β-N-Acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. *J Biol Chem*. 286:37483–37495.